

---

# A comparative study of six European databases of medically oriented Web resources

By Francisca Abad García, MD, MLIS, PhD  
abad@uv.es  
Professor of Medical Information Science  
Faculty of Medicine

Aurora González Teruel, MLIS, PhD  
agonzal@uv.es  
Professor of Library and Information Science

López Piñero Institute for the History of Science and Information Science (CSIC-UV)  
University of Valencia  
Faculty of Medicine  
Blasco Ibáñez, 15  
46010 Valencia  
Spain

Patricia Bayo Calduch, MD  
patriciabayo@comv.es  
Resident Medical Doctor

Doctor Peset University Hospital  
Valencia  
Spain

Rosa de Ramón Frias, MD  
romdera@alumni.uv.es  
Medical Doctor

La Fe University Hospital  
Valencia  
Spain

Lourdes Castillo Blasco, MLIS  
M.Lourdes.Castillo@uv.es  
Lecturer in Information Science

López Piñero Institute for the History of Science and Information Science (CSIC-UV)  
University of Valencia  
Faculty of Medicine  
Blasco Ibáñez, 15  
46010 Valencia  
Spain

---

**Objectives:** The paper describes six European medically oriented databases of Web resources, pertaining to five quality-controlled subject gateways, and compares their performance.

**Method:** The characteristics, coverage, procedure for selecting Web resources, record structure, searching possibilities, and existence of user assistance were described for each database. Performance indicators for each database were obtained by means of searches carried out using the key words, "myocardial infarction."

**Results:** Most of the databases originated in the 1990s in an academic or library context and include all types of Web resources of an international nature. Five databases use Medical Subject Headings. The number of fields per record varies between three and nineteen. The

language of the search interfaces is mostly English, and some of them allow searches in other languages. In some databases, the search can be extended to Pubmed. Organizing Medical Networked Information, Catalogue et Index des Sites Médicaux Francophones, and Diseases, Disorders and Related Topics produced the best results.

**Conclusions:** The usefulness of these databases as quick reference resources is clear. In addition, their lack of content overlap means that, for the user, they complement each other. Their continued survival faces three challenges: the instability of the Internet, maintenance costs, and lack of use in spite of their potential usefulness.

## INTRODUCTION

The Internet has brought about an unprecedented revolution in the capacity to access and disseminate information. Although information has never before been so accessible, the nature of the Internet does not always facilitate finding that information. Statistics about the size of the Web fluctuate but always agree that it is virtually unlimited. The quantity of 3 billion Web documents has been quoted [1] (of which some 2% correspond to pages concerned with health) a quantity that doubles, according to current assessments, every 173 days [2].

Diversity is a consistent characteristic of Web documents: they disseminate material from all over the world, are written in any language, have diverse content, and are directed to different audiences. Thus, the same subject can be treated according to a desire to inform, an orientation to the general public, or a presentation from a professional perspective. Another cause for concern is the diverse quality of the contents of the Web. Documents that have been revised by peer professionals (scientists, for example), coexist with others that have received no checking, as anyone can disseminate information on the Internet, taking on the role of author, editor, and distributor at the same time [3, 4].

Several initiatives are aimed at controlling the problem of quality in Web documents on health issues. Among them are [3–6]:

- strategies aimed at educating users in applying criteria to evaluate the credibility of the Websites [7–9]
- initiatives aimed at encouraging producers of Websites to follow principles based on ethical codes such as those proposed by the Health On the Net Foundation (HON) [8] or the e-health Code of Ethics of the Internet Health Coalition [9] on the design of Web pages, as well as the incorporation of metadata into pages that facilitate identifying the origin of the data and the authorship of the contents
- projects based on institutions or agencies acting as filters by assigning awards or seals of approval to indicate that a site has met quality standards [10, 11]
- an alternative proposed by the European project MedCERTAIN [6, 12], which is based on the incorporation of a filtering tool that works by supplying the user's computer with the appropriate software for filtering information

The aim of all of these is to deal with the problem of quality. They do not, however, face the problem of determining what quality information exists on the Internet for professional use and ways to access it to meet specific needs. Both of these objectives are addressed, for example, by quality-controlled subject gateways. These are defined by Koch as,

Internet-services which apply a rich set of quality measures to support systematic resource discovery. Considerable manual effort is used to secure a selection of resources, which meet quality criteria, and to display a rich description of these resources with standards-based metadata. Regular checking and updating ensure good collection management. A main goal is to provide a high quality of subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing. [13]

Gateways store the resource description in a database and assign one or more searching tools. These databases consist of procedures for finding, filtering, describing, cataloging, and maintaining a set of Web resources on a particular subject (in this case, medicine). They are enormously valuable tools for health professionals, users who are traditionally very short of time and are sometimes not computer experts [14, 15], as well as for health sciences librarians.

## OBJECTIVES

This paper aims to: (1) to describe the characteristics of six European databases of medically oriented Web resources, pertaining to five quality-controlled subject gateways; (2) to compare their function by obtaining performance indicators; and (3) to pinpoint the problems and challenges of these kind of sources.

## METHODOLOGY

The databases have the following elements in common: they are free access, they carry out manual selection (totally or partially) of the resources, and they all, except one, use Medical Subject Headings (MeSH) for their indexing. These sources were located by means of checking relevant publications and the Internet. Given the nature of the medium, however, it cannot be claimed that they are the only ones in existence, although they are almost certainly the most representative.

The databases pertain to the following gateways:

- Computer Aid Learning (CAL) Reviews [16]
- Catalogue et Index des Sites Médicaux Franco-phones (CISMeF) [17]
- Diseases, Disorders and Related Topics (DDRT) [18]
- HON foundation [19]: Medical Document Hunter (MedHunt) [20] and the Multilingual and Intelligent Search Tool Integrating Heterogeneous Web Resources (HONSelect) search engine [21]
- Organizing Medical Networked Information (OMNI) [22]

In the case of HON, the MedHunt database [20] and HONSelect search engine [21] are described separately. The elements described for each database are:

- Characteristics and coverage: name, uniform resource locator (URL), gateway name, institution, country where produced, date of creation, context, source of financing, audience, language or origin of the resources, number of resources, and type of resource
- Collection development and resource description: specific mention of the selection criteria, existence of an editorial committee, staff dedicated to carrying out the selection and/or cataloging of the resources, the language used for indexing, fields per record, number of Dublin Core elements [23] included, and mention of the method of checking the links
- Access to the collection: interface language, search language, types of indexes, types of searches, and possibility of extending the search to PubMed and to other databases
- Others: existence of guides or tutorials, publications about the resource, or awards received

The information was obtained by means of direct checking and by any publications that existed about them. The last check was carried out in February 2005. To obtain the performance indicators for each database, the authors carried out a search using the simple search mode and the alphabetical index. The keywords used were "Myocardial infarction" or "Infarctus du myocarde," depending on whether the search was carried out in English or in French. The indicators were:

- the total number of resources retrieved
- the number of different resources retrieved
- the number of resources retrieved with working links (named active resources)
- the number of relevant active resources
- the proportion of relevant active resources among the retrieved active resource (precision ratio) [24]
- the number of common resources retrieved when using different search procedures (overlap) [25]

The relevance of the resources was evaluated independently by two medical doctors, whose level of agreement on meeting the relevant criteria was 87.5%. They resolved their disagreements to reach the final evaluation of the resources. The resources have been classified in three categories:

1. relevant: the contents of the resource are concerned with myocardial infarction (prevention, epidemiology, diagnosis, pathology, treatment, rehabilitation, etc.)
2. partially relevant: the contents of the resource, al-

though related to some extent, do not deal specifically with the subject of infarction

3. not relevant: the contents of the resource are concerned with other medical subjects

The results of simple search mode searches in these databases were also compared using the relative recall ratio (the proportion of active relevant resources retrieved by each database compared with the number of unique relevant active resources retrieved by the set of databases) [26] and the global overlap between databases (the number of common relevant active resources retrieved from different databases) [25].

## FINDINGS

### Characteristics and coverage of the resources

The databases analyzed here are from various different European countries (Table 1). The oldest (DDRT) dates from 1993 and the most recent (HONSelect) from 1999. Most of them were created in an academic or librarianship context, except for the HON databases, and are publicly financed. Their intended users are mainly specialists (health professionals researchers, teachers, etc.), although patients are not excluded.

The origin of the resources is international. With regard to language, resources in English clearly predominate, with the exception of CISMeF, for which the resources are mainly in French. All types of resources are included, irrespective of the kind of document or content orientation (academic, clinical, or patient, etc.) or whether it is free to the user or not. The pedagogical nature of the CAL Reviews should be noted, as well as the clinical practice nature of the contents of DDRT. The number of resources is not always known.

### Collection development and resource description

Table 2 highlights the fact that only three of the gateways (CAL, CISMeF, and OMNI) give information about the criteria used for selecting and evaluating the resources. Only in the case of CISMeF is the selection and inclusion of the Web resources based on an editorial committee, although the remainder mention a group of assessors. Cataloging is carried out equally by librarians and medical doctors, or even students, as in the case of CAL Reviews.

With regard to selection and cataloging, special mention be made of the two HON databases (MedHunt and HONSelect). MedHunt is a database that unites resources automatically selected and cataloged by a powerful search engine known as the Multi Agent Retrieval Vagabond on Information Network (MARVIN) and manually by HON staff. MARVIN is a search engine for finding medical Web resources. It makes a distinction between these and nonmedical resources by means of algorithms that compare the vocabulary of the Website with MARVIN's own glossary of medical terms. A relevance score is assigned to the Website according to the frequency of the appearance of specific terms and according to their position on the page [35]. On the other hand, the selection and cataloging criteria for the resources indexed by HON members

**Table 1**  
Description of the characteristics and coverage of the databases

Name:	CAL Reviews	CISMeF	DDRT	MedHunt	HONSelect	OMNI
Uniform resource locator (URL) Gateway name Institution	<a href="http://axis.dbcu.cam.ac.uk/calreviews/">http://axis.dbcu.cam.ac.uk/calreviews/</a> CAL Centre for Medical Interactive Technology and the Clinical and Biomedical Computer Unit of Cambridge University	<a href="http://www.chu-rouen.fr/cismef/">http://www.chu-rouen.fr/cismef/</a> CISMEF Centre Hospitalier Universitaire de Rouen	<a href="http://www.mic.ki.se/Diseases/DDRT">http://www.mic.ki.se/Diseases/DDRT</a> Karolinska Institutet University Library	<a href="http://www.hon.ch/MedHunt">http://www.hon.ch/MedHunt</a> HON Health On the Net Foundation	<a href="http://www.hon.ch/HONSelect/">http://www.hon.ch/HONSelect/</a> HON Health On the Net Foundation	<a href="http://omni.ac.uk">http://omni.ac.uk</a> OMNI Partnership between the University of Nottingham and the Medical Library, the Royal Free, University College London Medical School, and the Royal Free Hampstead NHS Trust
Country Date of creation	United Kingdom 1997 [27]	France 1995 [28]	Sweden 1993 in Gopher and since 1995 by means of Web access [29] Librarian Not known	Switzerland 1996 [30]	Switzerland 1999 [31]	United Kingdom 1995 [29]
Context Source of finance	Academic Service Increment for Teaching (SIFT)	Academic/librarian Centre Hospitalier Universitaire de Rouen	Not known	Medical and professional The Geneva Ministry of Health, the Geneva University Hospital, and the Swiss Institute of Bioinformatics	Medical and professional The Geneva Ministry of Health, the Geneva University Hospital, and the Swiss Institute of Bioinformatics	Academic/librarian Joint Information Systems Committee (JISC) through the Resource Discovery Network < <a href="http://www.rdn.ac.uk">http://www.rdn.ac.uk</a> >
Audience	Students and teachers	Students, researchers, teachers, health professionals, and patients	Professionals and patients	Professionals, patients, and health service providers	Professionals, patients, and services providers	Students, researchers, teachers, medical doctors, and patients
Language or geographical origin of the resources	Mainly in English	Francophone (70% French, 16% Canadians, etc.) and the remainder in other languages	Mainly in English; separate access for resources in Swedish [32]	English, French, German, Spanish, and Portuguese	English, French, German, Spanish, and Portuguese	Mainly in English
Number of re-sources	Not known	≥ 15,000 [28]	16,000 resources [33]	Not known	Not known	≥ 8,672 [34]
Type of resource	Academic oriented; free and commercial	All types; free and commercial	Clinically oriented; free and commercial	All types; free and commercial	All types; free and commercial	All types; free and commercial

**Table 2**  
Collection development and resource description

	Name:	CAL Reviews	CISMeF	DDRT	MedHunt	HONSelect	OMNI
Specific mention of the selection criteria	Yes < <a href="http://axis.cbcu.cam.ac.uk/calreviews/guide.htm">http://axis.cbcu.cam.ac.uk/calreviews/guide.htm</a> >	Yes < <a href="http://www.chu-rouen.fr/netscoring/">http://www.chu-rouen.fr/netscoring/</a> >	No	No	No: automatic selection by MARVIN and manual selection by HON staff	No: automatic selection from MedHunt	Yes < <a href="http://biome.ac.uk/guidelines/collection/">http://biome.ac.uk/guidelines/collection/</a> >
Editorial committee	No (Cambridge clinical teachers act as moderators in resources review)	Yes (membership not specified) [36]	No	No	No (medical assessment group)	No (medical assessment group)	No (medical assessment group)
Who does the cataloguing	Students and clinical school staff	Medical librarians	Medical librarians?	Medical librarians?	Automatic (MARVIN); doctors	Automatic; doctors	Information specialist
Indexing language	Not known						
Fields per record	Title, URL, author/producer, summary, suitability for the Cambridge clinical courses, reviewer, strong points, weak points, use purpose, course areas, year, phase, review summary, product details, keywords, support, price	MeSH, National Library of Medicine (NLM) classification scheme	MeSH	MeSH	Purpose-built medical dictionary of 20,000 words (in 5 languages) plus the 33,000 MeSH terms	Purpose-built medical dictionary of 20,000 words (in 5 languages) plus 33,000 MeSH terms	MeSH
		Title, author or creator, editor, publisher, identifier, resource description, audience, language, country, date of publication, subject, keywords, kind of document, resource type, access, format, URL, financing, rights	Title, source, country	Title, source, country	■ Manual: title, URL, resource description, type, location, keywords, language, registration and HON review, subsequent HON monitoring, visiting date by MARVIN, relevance query and adhesion to the HON code	Title, URL MeSH terms, date of MARVIN visit	Title, resource description, MeSH descriptors
Fields per record	17	19	3	3	Manual: 13 Automatic: 7	3	3
Number of Dublin Core items	5	11	3	3	Automatic: 7	3	3
Link checking	Checked automatically on a weekly basis	The checking of links is automatic and is carried out monthly; broken links are found manually	Procedure not known	Procedure not known	Procedure not known	Procedure not known	The links are automatically and permanently revised; the contents of each record are revised by its indexer every 9 months



are not indicated. With regard to HONSelect, the Web resources are obtained from MedHunt by means of an automated selection based on the relevance of the document to a specific MeSH term [31].

The indexing language used in all of the databases is MeSH, except in the case of CAL Reviews. The number of fields per record is very variable, from three for DDRT and OMNI to nineteen for CISMef. CISMef has the greatest number of fields that correspond to the Dublin Core recommendations. Only three gateways explain their procedure for checking links.

### Access to the resources

English is the dominant language, both with regard to the Web pages themselves and to the search interfaces, except CISMef. It should be noted that some of the databases enable searches in more than one language, one of which is always English (Table 3). Some databases have special features that should be pointed out. For example, MedHunt enables searches in five different languages (English, French, German, Spanish, and Portuguese), but the response depends on the language used for the search. When the database is consulted in Spanish, it does not find all of the resources for this keyword, but only resources that are in Spanish. Furthermore, if the resource in question is indexed by HON, the description is usually in Spanish. In the case of CISMef, however, its search engine, Doc'CISMef <<http://doccismef.chu-rouen.fr/servlets/Simple/>>, enables searches equally in English and in French, without affecting the results.

All of the databases facilitate access to the resources by means of various kinds of indexes, except for those that depend on the HON gateway. Some databases add options for viewing the resources, such as the ability to consult the most recent resources added to the database (CISMef, CAL Reviews, and OMNI) or to view only resources that are aimed at a particular public, resources that refer to a specific type of document (CISMef), or resources that are written in a particular language (DDRT).

The most common search method is the simple search mode option, and some characteristics with regard to searching should be pointed out. For example, the simple search mode for MedHunt presents the results listed according to the following criteria: resources that are in accordance with the HON code of conduct, resources that are indexed by HON staff, and resources that are selected automatically by MARVIN. In this way, the user can either view all of the resources or only those that meet particular criteria.

A particular characteristic of the HONSelect database is that, using the MeSH terms, it integrates the search results with the search results of five other databases: Hon Media (more than 1,700 medical images), MEDLINE, News Page (news from newspapers), and Clinicaltrials.Gov (clinical trials). In this way, it is possible to obtain various lists of resources in one search.

OMNI is similar, as it deals with one of the databases of the BIOME hub, created in 1998. BIOME also includes the following databases: NMAP (nursing,

midwifery, and allied professions), BioResearch (biological and biomedical research), VetGate (animal health and veterinary science), and Agrifor (agriculture, food, and forestry), and it hosts the MedHist (history of medicine), BioethicsWeb (bioethics), and Psicom (science and technology) databases from the Wellcome Trust. This search method enables an advanced search to be carried out only in OMNI or extended to any of the previously cited databases.

Three databases (CISMef, DDRT, and HONSelect) provide the option of extending the search to the bibliographical database PubMed, as the indexing language used is MeSH. DDRT, in addition, allows the search to be extended to other bibliographic databases and library catalogs, and CISMef allows the search to be extended to other NLM resources and to OMNI.

With regard to information provided by the Website of each database on the origin of the data, selection criteria, methodology used, and a user guide to the resource, only two of the analyzed databases, CISMef and OMNI, provided detailed information.

### THE SEARCH RESULTS

The results of the searches for each of the databases and the search method applied are described below and summarized in Table 4.

#### Computer Aid Learning (CAL) Reviews

The data showed that the same results were obtained if the search was by title or by title and description together. A total of four resources were found, of which three were repeated, because this database showed the resources according to the specialties they were assigned to. Of the two unique resources, only one had an active link and its contents were considered relevant.

#### Catalogue et Index des Sites Médicaux Francophones (CISMef)

The CISMef simple search mode option was carried out in French, using the MeSH term, "Infarctus du Myocarde." For this search, 53 resources were found, of which none were repeated. A total of 44 resources (61.4% of the total) had an active link, and, of these, 22 were considered relevant to the search, representing 48.96% precision. Use of the alphabetical index enabled visualization of all of the resources (the same 53 obtained by means of the simple search mode option) or selection of only the main ones (18 resources) or those whose contents are recommendations (14 resources), teaching documents (29 resources), or information for patients (7 resources). As expected, all of the resources found using the alphabetical index were the same as those found in the simple search mode.

#### Diseases, Disorders and Related Topics (DDRT)

The results demonstrated different functioning methods in the page search and link search facilities. Neither of them produced repeated resources. The page

**Table 3**  
Description of means of access to the resources and to the information provided by the gateway

Name:	CAL Reviews	CISMeF	DDRT	MedHunt	HONSelect	OMNI
Interface language	English	French	English	English and French	English	English
Search language	English	French and English	English	English, French, German, Spanish, and Portuguese	English, French, German, Spanish, and Portuguese	English
Access to the resources by means of indexes	Indexes by specialty, category	Alphabetical index, subject index, and by type of resource	Alphabetical index of illnesses	No index available	No index available	Index by categories of the NLM classification
Other possibilities of viewing the resources	Most recent resources (What's New); editor's choice	Most recent resources; separate access to information for patients, recommendations, consensus documents, and teaching resources	Separate access to resources in Swedish < <a href="http://www.mic.ki.se/Diseases/sweptat.html">http://www.mic.ki.se/Diseases/sweptat.html</a> >	No other possibilities	No other possibilities	Latest additions
Access to the resources by searches	Simple search mode option, by title, by title and description	Simple search mode option; advanced search	Page search; link search	Simple search mode option; the search can be refined or limited; the results enable separate access to resources that meet the HON code, to those cataloged manually, and to those found by MARVIN; the language of the resources is the same as that of the search	Simple search mode option in normal language and by descriptors; the results are combined with searches in other databases: HON Media, MEDLINE, News Page, and Clinicaltrials.gov	Simple and advanced searches; the advanced search allows the joint use of OMNI with other databases of the BIOME portal (NMAP, BioResearch, VetGate, Agri-for); the advanced search enables limiting the search results to UK resources only
Search can be extended to PubMed	No	Yes	Yes	No	Yes	No
Search can be extended to other databases	No	NLM resources, OMNI	Other bibliographical databases, library catalogs	No	No	No
User guide or tutorials	No	Yes [37]	Not available	Not available	Not available	Yes [38]
Information about the resource	Yes, average	Yes, a lot	Yes, little	Yes, little	Yes, little	Yes, a lot
List of publications	No	Yes	No	Yes	Yes	No
Awards	No mention	Yes	No mention	Yes (to HON)	Yes (to HON)	No mention

**Table 4**  
Results of the searches, by database and by search method

Name:	CAL Reviews	CISMeF	DDRT	MedHunt	HONSelect	OMNI
<b>Search</b>	<b>A) By title</b>	<b>A) Simple search mode option</b>	<b>A) Page search</b>	<b>A) With HON code</b>	<b>D) Search (results in English)</b>	<b>A) Simple search mode option</b>
Number of re-sources re-trrieved	4	53	14	11	22	32
Number of different resources	2	53	14	11	22	32
Number of active resources	1	45	14	10	22	32
Number of relevant active resources	Relevant: 1	Relevant: 22; partially relevant: 17; not relevant: 8	Relevant: 12; partially relevant: 2; not relevant: 0	Relevant: 1; partially relevant: 4; not relevant: 5	Relevant: 12; partially relevant: 7; not relevant: 3	Relevant: 29; partially relevant: 3; not relevant: 0
Precision ratio	100.0%	48.9 %	85.7%	10.0%	54.5%	90.6%
	<b>B) Title and description</b>		<b>B) Link search</b>	<b>B) Described by HON</b>		
Number of re-sources re-trrieved	4		7	20		
Number of different resources	2		7	20		
Number active re-sources	1		7	18		
Number of relevant active resources	Relevant: 1		Relevant: 7	Relevant: 3; partially relevant: 6; not relevant: 9		
Precision ratio	100.0%		100.0%	16.7%		
Overlap	$A \cap B = 4$		$A \cap B = 7$	$A \cap B = 0$		
				<b>C) MARVIN (*the first 20 out of 1,188)</b>		
Number of re-sources re-trrieved				20		
Number of different resources				20		
Number of active resources				20		
Number of relevant active resources				Relevant: 13; partially relevant: 2; not relevant: 5		
Precision ratio				68.4%		
Overlap				$A \cap B \cap C = 0$ $A \cap D = 0; B \cap D = 0; C \cap D = 5$		
<b>Navigation</b>		<b>B) Alphabetical index MeSH terms</b>	<b>C) Alphabetical index MeSH terms</b>			<b>B) Alphabetical index MeSH terms</b>
Number of re-sources re-trrieved		53	14			15
Number of different resources		53	14			15
Number of active resources		45	14			15
Number of relevant active resources		Relevant: 22	Relevant: 12			Relevant: 15
Precision ratio		48.8%	48.7%			
Overlap		$A \cap B = 53$	$A \cap C = 14$ $B \cap C = 7$			$A \cap B = 15$

search found fourteen resources, all of them with active links. Twelve were considered relevant. The results of the link search were fewer, only seven resources that were active and relevant. The resources found by means of the link search were included in the fourteen obtained by means of the page search. The results obtained using the MeSH term, "Myocardial Infarction," in the alphabetical index were the same as those found using the page search. It can be concluded from

this that, when searching, the page search looks at title and description, while the link search only looks at the title.

#### Medical Document Hunter (MedHunt)

MedHunt produced the following results: 11 resources conformed to the HON code, 20 resources had been manually described by HON, and 1,188 resources were



**Table 5**  
Comparison between resources resulting from a simple search mode option

Database (K kind of search)	Number of different resources (a)	Number of active resources (b)	% Active resources (b/a × 100)	Number of relevant active resources (d)	Precision ratio (d/b × 100)	Relative recall ratio (d/f × 100)
CAL Reviews (title & description)	2	1	50.0	1	100.0	1.3
CISMeF (simple search mode option)	53	45	84.9	22	48.9	29.3
DDRT (Page search)	14	14	100.0	12	85.7	16.0
MedHunt (resources visited and described by HON)	19	17	89.5	3	16.7	4.0
HonSelect (simple search mode option)	22	22	100.0	12	54.5	16.0
OMNI (simple search mode option)	32	32	100.0	29	90.6	38.7
Total (e)	142	130	91.5	78	60.0	
Total without repetitions (f)	139	127		75	59.1	
Global overlap (f – e/e × 100)	2.2%	2.3%		3.8%		

found by MARVIN. As it was not possible to evaluate these 1,188 resources, the authors chose the same number of retrieved resources that were manually described by HON, and the first 20 were, therefore, evaluated. The findings were evaluated according to 3 categories. Table 4 shows that there were no repeated resources in any of the 3 categories and all, except 1, were active links. The number of relevant links found was very small: 1 in the case of the resources that accorded with the HON code, and 3 in the resources indexed by HON members. The largest number of relevant resources was obtained when the first 20 resources found by MARVIN were analyzed: 13 resources, indicating a 68.4% precision. No resources were common to 2 or more categories.

#### Multilingual and Intelligent Search Tool Integrating Heterogeneous Web Resources (HONSelect)

This database provided twenty-two different resources. Twelve of these were considered relevant (54.5% precision). When the resources retrieved by HONSelect were compared with those resources retrieved by MedHunt, only five common links were found, all of them found automatically by MARVIN.

#### Organizing Medical Networked Information (OMNI)

OMNI's simple search mode option provided thirty-two resources without repetition and with active links, of which thirty (90.6%) were relevant. When the alphabetical index of MeSH terms was used to search for resources, fifteen active resources were found, all of them relevant. Comparison of the resources obtained with the two search methods showed that those obtained by means of the alphabetical index were all included in the results obtained through the simple search mode option.

#### RESULTS OF THE COMPARISON BETWEEN DATABASES

Table 5 shows the comparison of the number of resources found in the various databases using the simple search mode. In total, the 6 databases together produced 142 resources, of which 3 were repeated, an overlap of 2.2%. The percentage of resources that were

active was very high for the resources taken together (91.5%) as well as for each database taken separately, as the value obtained for CAL Reviews was negligible given that it only had 2 resources.

Precision was 60.0% for the resources taken together. This measure varied considerably from one database to another. The OMNI, DDRT, and HONSelect databases had the highest precision ratio. The OMNI and CISMeF databases had the highest relative recall ratio.

#### DISCUSSION

"Quality-controlled subject gateway" is the term currently used to indicate services that integrate databases of Web resources such as those described here [13]. This term has not been used here, however, because the authors are more concerned with describing the characteristics of the databases as searching tools than in describing the gateways in their totality.

Among the general characteristics of these databases, it should be noted that their development results from the joint work of various kinds of professionals, such as librarians, information scientists, and health professionals. The active role of librarians should be noted, as they have gone beyond being merely providers of products created by others to become leaders in the creation of new resources that can provide answers to problems and needs in the context of the Internet.

Another characteristic to be considered is the lack of structural standardization of these sources. Thus, while all of the databases except CAL Reviews, incorporate the MeSH thesaurus [39], a tool that doctors and librarians are familiar with, to assign descriptors, none of the databases adopt the entire list of fifteen Dublin Core elements [23], which are recommended as a standard model for the description of Web resources. Similarly, the frequency with which information on criteria, methodology, and instructions on the use of the databases is either absent or difficult to find, which is also an indication of this lack of standardization. Ignorance of some characteristics of these databases could limit their usefulness or, in some cases, disorient the user.

With regard to their utility as sources of information, the fact that relevant resources are retrieved to respond to a particular illness shows that they are indeed useful as quick reference tools. It should also be

pointed out that the fact that the links were generally active indicates that the resources are, on the whole, well maintained.

In spite of the fact that the number of resources found in the searches was not great (given the quantity of resources contained in the databases and the specific nature of the subject, this was to be expected) and that it could not be claimed that these databases would produce the same quantity of results if different key words were used, the value of the indicators was of great use in advancing knowledge of the way these sources functioned. Thus, in the case of DDRT, it has enabled understanding of the behavior of the database when the page search method (i.e., title and description) or the link search (i.e., title) is used, which is not otherwise made clear anywhere.

In the case of CISMef, the results make clear that using the simple search mode option is equivalent to using the alphabetical index of MeSH terms and that the advantage of the latter option is that it enables finding resources according to the nature of their contents. Similarly, the statistics for precision with regard to this database leads one to assume the use of exhaustive criteria in the indexing.

The characteristics of OMNI seem to be quite different. In this database, when the alphabetical index of MeSH terms is used, the findings are fewer than when the simple search mode is used. This result indicates the use of more precise indexing criteria. Thus, to obtain the most exhaustive findings, the simple search mode option is recommended.

The most noteworthy example is that of the behavior of the HON gateway databases. First, considering MedHunt, the resources found automatically by MARVIN clearly have greater relevance than the results indexed manually by HON staff. Second, although it is claimed that HONSelect obtains its results from MedHunt [40], the overlap between the two databases is low (nearly zero), which, from a practical point of view, leads to considering them as independent of each other and implies both sources should be used.

When all the sources and their functioning are considered in their totality, the small amount of overlap among the various databases is significant. This fact also leads to the conclusions, first, that each database functions independently and that none of them functions as the finding resource for others and, second, that the sources are complementary and that, if one wishes to carry out the widest possible search, all of the databases should be used. The complementary nature of the databases is indicated, for example, by the fact that CISMef has recently introduced the possibility of extending a search to OMNI. It is presumably for this reason that the possibility of creating a megagateway of health gateways sharing the same thesaurus (MeSH)—which includes CISMef, DDRT, HON, and OMNI—has recently been proposed [41].

It is clear, therefore, that all of the studied sources have strong and weak points with regard to their characteristics and functioning. Thus, to view each of them in a more global context, to present an overall sum-

mary of each of them, and to limit space, this information is set out in tabular form (Table 6). From this table, it is clear that the OMNI, CISMef, and DDRT databases, which are in fact the oldest, are the databases with the most favorable assessment, although this does not doubt the usefulness of the others.

In spite of the obvious advantages in having databases of Web resources available, setting up this kind of product implies facing three significant challenges: the instability of the Internet itself, the costs of the management and maintenance of these sources, and the use that professionals will make of this kind of source.

The instability of the Internet is an issue that concerns mainly the survival of the resources and the lack of continuity of their locations. Thus, while documents that formed part of traditional bibliographic databases more than thirty years ago can still be found today with their contents unchanged, on the contrary, with the Internet, it is impossible to know if resources included in databases yesterday will still be available tomorrow in the same place or in a new location or even if they will have disappeared completely. A study published by Koelher in 1999, based on monitoring a sample of 361 Websites, has found that the percentage of pages active after 214 weeks of observation was no greater than 35% [42]. Moreover, resources on the Internet, far from being static, are being continually updated, and what is seen is no more than a snapshot, a scene from a film, the end of which cannot currently be predicted [43].

Several possible deductions follow from the two factors mentioned above. The first is the possibility that documents containing information to support statements, hypotheses, interpretations, or conclusions related to current research will disappear [44]. This is a much debated issue and one that should be borne in mind, although its implications go beyond the scope of this paper. The second factor is related to the management and maintenance of this type of database. The survival and movement of their resources and, consequently, the need for constant updating imply an extra workload and an increase in financial cost. To this workload and cost has to be added the need for descriptions of the contents to agree with subsequent updating of the pages, which implies taking the time to revise the contents periodically and re-index the material.

Maintenance costs, both intellectual and financial, as well as the need for multiskilled teams, are probably one of the most important reasons why these databases are so few and have an uncertain future. The described databases are financed by various types of public funds, though this is not necessarily a guarantee of their survival. In 2003, lack of financing was the main reason for the disappearance of CliniWeb <<http://www.ohsu.edu/clinWeb/>>, which was an important catalog of health and biomedical information on the Web created in 1995. A possible alternative is the commercialization of this type of product, although that means that access to them is no longer free. Examples

**Table 6**  
Qualitative assessment of the database described

Name:	Positive aspects	Negative aspects
CAL Reviews	<ul style="list-style-type: none"> <li>■ Teaching oriented</li> <li>■ Easy to use</li> <li>■ Abundant information on each item found</li> </ul>	<ul style="list-style-type: none"> <li>■ Very few resources found</li> </ul>
CISMeF	<ul style="list-style-type: none"> <li>■ Various means of accessing the resources, versatile</li> <li>■ Unique with regard to geographical and language coverage</li> <li>■ Exhaustive information about its creation, characteristics, and criteria for selecting and including the resources</li> <li>■ High level of standardization</li> <li>■ The description available for each resource is very professional and complete</li> <li>■ Extension of searches to PubMed and other databases</li> <li>■ Extension of searches to other Web resources databases</li> <li>■ Awards</li> </ul>	<ul style="list-style-type: none"> <li>■ Failures on precision may be due to exhaustive indexing</li> </ul>
DDRT	<ul style="list-style-type: none"> <li>■ Intuitive interface</li> <li>■ Clinical decision making oriented</li> <li>■ Precise searching</li> </ul>	<ul style="list-style-type: none"> <li>■ Summary description of each item found (and so it could be considered a catalog)</li> <li>■ Little information about its characteristics and functioning method</li> <li>■ The absence of a user guide may reduce the search efficacy</li> </ul>
MedHunt	<ul style="list-style-type: none"> <li>■ Intuitive interface</li> <li>■ Unique search facilities, versatile</li> <li>■ Powerful search engine</li> <li>■ Possibility of searching in other languages</li> </ul>	<ul style="list-style-type: none"> <li>■ Low level of precision in items indexed manually</li> <li>■ Unknown selection criteria for the resources entered manually</li> <li>■ The absence of a user guide may reduce the search efficacy</li> </ul>
HONSelect	<ul style="list-style-type: none"> <li>■ Simultaneous searches in five different databases</li> <li>■ Replies in several languages</li> </ul>	<ul style="list-style-type: none"> <li>■ Low level of precision</li> <li>■ The absence of a user guide may reduce the search efficacy</li> <li>■ Its relationship to MedHunt is not clear</li> </ul>
OMNI	<ul style="list-style-type: none"> <li>■ Simple and easy-to-use design</li> <li>■ Exhaustive information about its creation, characteristics, and criteria for selection and inclusion of resources</li> <li>■ The advanced search enables searching simultaneously in various BIOME databases</li> <li>■ The description available for each resource is very professional and complete</li> <li>■ High level of precision and recall</li> </ul>	

of this are the North American databases Medical Matrix <<http://www.medmatrix.org/reg/login.asp>>, access to which is restricted to subscribers, and Medical Word Search <<http://www.mwsearch.com>> or the Spanish QReME subsidized by the Colegio Oficial de Médicos de Valencia (COMV), the use of which is restricted to its members [45].

Commercializing this type of product, therefore, leads to the question, "Will the users appreciate the added value of the filtering, description, cataloging, and possibility of selective searches facilitated by this kind of resource to the extent of being prepared to pay for it, taking into account that they could access Web resources freely without an intermediary?"

Another threat to the survival of this kind of resource is lack of use, because, in spite of their potential usefulness or, indeed, necessity, they are little known by medical doctors and librarians, and, if they are not used, they may be condemned to disappear, as in the case of the BioSites database [46]. The risk of this happening would be less if these resources become part of the collection of resources medical doctors and health sciences librarians normally use and if these users were trained in their use as though they were simply another standard resource.

## CONCLUSIONS

■ Databases of Web resources are still scarce products but are of recognized usefulness in facing the

challenges raised in the framework of Internet retrieval.

■ The described databases have the potential to be useful as quick reference sources. The small amount of overlap between them indicates that the sources are complementary.

■ Of the databases described, OMNI, CISMeF, and DDRT were the most favorably assessed.

■ These kinds of resources face three significant challenges: the instability of the Internet, the costs of their management and maintenance, and the amount of use the databases receive.

## REFERENCES

1. POWELL J, CLARKE A. The WWW of the World Wide Web: who, what, and why? *J Med Internet Res* [serial online]. 2002 Jan-Mar;4(1):e4. [cited 14 Mar 2005]. <<http://www.jmir.org/2002/1/e4/>>.
2. MORENO ROMERO J. Pasado, presente y futuro de la recuperación de información en Internet. [Web document]. Paper presented at: Congreso Internacional sobre Retos de la Alfabetización Tecnológica en un mundo de red; Cáceres, 30 Nov-2 Dec 2000. [rev. 22 Dec 2002]. Cited in: Rodríguez Camiño R. Motores de búsqueda sobre salud en Internet.aciMED 2003 [serial online]. 11(5). [cited 14 Mar 2005] <[http://bvs.sld.cu/revistas/aci/vol11\\_5\\_03/aci02503.htm](http://bvs.sld.cu/revistas/aci/vol11_5_03/aci02503.htm)>.
3. SILBERG WM, LUNDBERG GD, MUSACCHIO RA. Assessing, controlling, and assuring the quality of medical information on the Internet: caveat lector et viewer—let the reader and viewer beware. *JAMA* 1997 Apr 16;277(15):1244-5.



4. HERSH WR, GORMAN PN, SACHEREK LS. Applicability and quality of information for answering clinical questions on the Web. *JAMA* 1998 Oct 21;280(15):1307-8.
5. PRICE SL, HERSH WR. Filtering Web pages for quality indicators: an empirical approach to finding high quality consumer health information on the World Wide Web. *Proc AMIA Symp* 1999;911-5.
6. EYSENBACH G, YIHUNE G, LAMPE K, CROSS P, BRICKLEY D. MedCERTAIN: quality management, certification and rating of health information on the Net. *Proc AMIA Symp* 2000; 230-4.
7. KIM P, ENG TR, DEERING MJ, MAXFIELD A. Published criteria for evaluating health related Web sites: review. *BMJ* 1999 Mar 6;318(7184):647-9.
8. HEALTH ON THE NET FOUNDATION. HON code of conduct for medical and health Web sites. [Web document]. Geneva, Switzerland: The Foundation, 1997. [cited 14 Mar 2005]. <<http://www.hon.ch/HONcode/Conduct.html>>.
9. INTERNET HEALTHCARE COALITION. E-health ethics initiative. [Web document]. Newtown, PA: The Coalition, 2003. [cited 14 Mar 2005]. <<http://www.ihealthcoalition.org/ethics/ethics.html>>.
10. JADAD AR, GAGLIARDI A. Rating health information on the Internet: navigating to knowledge or to Babel? *JAMA* 1998 Feb;279(8):611-4.
11. GAGLIARDI A, JADAD AR. Examination of instruments used to rate quality of health information on the Internet: chronicle of a voyage with an unclear destination. *BMJ* 2002 Mar;324(7337):569-73.
12. EYSENBACH G, DIEPGEN TL. Labeling and filtering of medical information on the Internet. *Methods Inf Med* 1999 Jun;38(2):80-8.
13. KOCH T. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Inform Rev* 2000; 24(1):24-34.
14. CULLEN RJ. In search of evidence: family practitioners' use of the Internet for clinical information. *J Med Libr Assoc* 2002 Oct;90(4):370-9.
15. ROGERS C. Physicians lack skills for computer CMC. *Am Acad Orthop Surg Bull* [serial online]. 2001 Apr;49(2). [cited 14 Mar 2005]. <<http://www.aaos.org/wordhtml/bulletin/apr01/cme1.htm>>.
16. CAMBRIDGE UNIVERSITY, CLINICAL & BIOMEDICAL COMPUTING UNIT. CAL reviews. [Web document]. Cambridge, UK: The Unit, 1997. [cited 14 Mar 2005]. <<http://axis.cbcu.cam.ac.uk/calreviews/>>.
17. CENTRE HOSPITALIER UNIVERSITAIRE DE ROUEN. CISMeF-Catalogue et index des sites médicaux francophones. [Web document]. Rouen, France: The Centre, 1995. [cited 14 Mar 2005]. <<http://www.chu-rouen.fr/cismef/>>.
18. KAROLINSKA INSTITUTET UNIVERSITY LIBRARY. DDRT: diseases, disorders and related topics. [Web document]. Stockholm, Sweden: The Library, 1995. [cited 26 Jul 2004]. <<http://www.mic.ki.se/Diseases/>>.
19. HEALTH ON THE NET FOUNDATION. HON's initiatives & services. [Web document]. Geneva, Switzerland: The Foundation, 2003. [cited 14 Mar 2005]. <<http://www.hon.ch/Project/>>.
20. HEALTH ON THE NET FOUNDATION. MedHunt: medical document hunter. [Web document]. Geneva, Switzerland: The Foundation, 1996. [cited 14 Mar 2005]. <<http://www.hon.ch/MedHunt/>>.
21. HEALTH ON THE NET FOUNDATION. HONSelect. [Web document]. Geneva, Switzerland: The Foundation, 1999. [cited 14 Mar 2005]. <<http://www.hon.ch/HONSelect/>>.
22. UNIVERSITY OF NOTTINGHAM GREENFIELD MEDICAL LIBRARY. OMNI (Organising Medical Networked Information). [Web document]. Nottingham, UK: The Library, 1995. [cited 14 Mar 2005]. <<http://omni.ac.uk>>.
23. DUBLIN CORE METADATA INITIATIVE. Dublin Core metadata element set, version 1.1: reference description. [Web document]. Dublin, OH: OCLC Research, 2003. [cited 14 Mar 2005]. <<http://dublincore.org/documents/dces/>>.
24. CLEVERDON C. The Cranfield tests on index language devices. *Aslib Proc* 1967 Jun;19(6):173-94.
25. GLUCK M. A review of journal coverage overlap with an extension to the definition of overlap. *J Am Soc Inf Sci* 1990 Jan;41(1):43-60.
26. LANCASTER FW. If you want to evaluate your library. 2nd ed. Champaign, IL: University of Illinois, Graduate School of Library and Information Science, 1993.
27. CLINICAL AND BIOMEDICAL COMPUTING UNIT, UNIVERSITY OF CAMBRIDGE. CAL reviews. [Web document]. Cambridge, UK: The Unit, 2003 [cited 14 Mar 2005]. <<http://www.cbcu.cam.ac.uk/cbcu/projects/calreviews.htm>>.
28. CISMEF TEAM. CISMeF: Catalog and index of French-language health Internet resources: a quality-controlled subject gateway. [Web document]. Rouen: CISMeF, 2002. [cited 14 Mar 2005]. <<http://www.chu-rouen.fr/cismef/cismefeng.html>>.
29. DARMONI SJ, THIRION B. Indexing the Web? a comparative study of three medical Web servers on the Internet: CliniWeb, Diseases, Disorders and Related Topics, Omni. [Web document]. Mednet 96, European Congress of the Internet in Medicine; Brighton, Royaume-Uni, Oct 1996:5-6. [cited 14 Mar 2005]. <<http://www.chu-rouen.fr/dsii/publi/mdntdl4.html>>.
30. HEALTH ON THE NET FOUNDATION. About Health On the Net Foundation. [Web document]. Geneva, Switzerland: The Foundation, 2003. [cited 14 Mar 2005]. <<http://www.hon.ch/Global/#15>>.
31. BOYER C, BAUJARD V, GRIESSER V, SCHERRER JR. HON-Select: a multilingual and intelligent search tool integrating heterogeneous Web resources. *Int J Med Inf* 2001 Dec;64(2-3):253-8.
32. KAROLINSKA INSTITUTET UNIVERSITY LIBRARY. Information om Sjukdomar och Medicin. [Web document]. Stockholm, Sweden: The Library, 1995. [cited 14 Mar 2005]. <<http://www.mic.ki.se/Diseases/swepat.html>>.
33. KAROLINSKA INSTITUTET UNIVERSITY LIBRARY. Biomedical links. [Web document]. Stockholm, Sweden: The Library, 2004. [cited 14 March 2005]. <[http://kib.ki.se/tools/medres/index\\_en.html](http://kib.ki.se/tools/medres/index_en.html)>.
34. UNIVERSITY OF NOTTINGHAM GREENFIELD MEDICAL LIBRARY. OMNI. what's new. [Web document]. Nottingham, UK: The Library, 2005. [cited 14 Mar 2005]. <<http://omni.ac.uk/whatsnew/>>.
35. HEALTH ON THE NET FOUNDATION. MARVIN: how does it work? [Web document]. Geneva, Switzerland: The Foundation, 2002. [cited 14 Mar 2005]. <<http://www.hon.ch/Project/Marvin-specificities.html>>.
36. DARMONI SJ, LEROY JP, BAUDIC F, DOUYERE M, PIOT J, THIRION B. CISMeF: a structured health resource guide. *Methods Inf Med* 2000 Mar;39(1):30-5.
37. CENTRE HOSPITALIER UNIVERSITAIRE DE ROUEN. Aide à la recherche. [Web document]. Rouen, France: The Centre, 2003. [cited 14 Mar 2005]. <<http://www.chu-rouen.fr/cismef/aide/indexaccueil.html>>.
38. UNIVERSITY OF NOTTINGHAM GREENFIELD MEDICAL LIBRARY. BIOME Search help. [Web document]. Nottingham, UK: The Library, 2004. [cited 14 Mar 2005]. <<http://biome.ac.uk/searchhelp.html>>.
39. NATIONAL LIBRARY OF MEDICINE. MeSH. [Web document]. Bethesda, MD: The Library, [1975]. [cited 14 Mar 2005]. <<http://www.nlm.nih.gov/mesh/MBrowser.html>>.
40. HEALTH ON THE NET FOUNDATION. HONSelect: a mul-

tilingual and intelligent search tool integrating heterogeneous Web resources. [Web document]. Geneva, Switzerland: The Foundation, 2002. [cited 14 Mar 2005]. <<http://www.hon.ch/Project/HONSelect.html>>.

41. THIRION B, DOUYÈRE M, SOUALMIA LF, DAHAMNA B, LEROY JP, DARMONI SJ. Metadata element sets in the CISMef quality-controlled health gateway. DC-2004, International Conference on Dublin Core and Metadata Applications; Shanghai, China; Oct 2004.

42. KOEHLER W. Digital libraries and World Wide Web sites and page persistence. *Inform Res* [serial online]. 1999 Jul;4(4). [cited 14 Mar 2005]. <<http://informationr.net/ir/4-4/paper60.html>>.

43. BROOKS TA. Web search: how the Web has changed information retrieval. *Inform Res* [serial online]. 2003 Apr;8(3).

[cited 14 Mar 2005]. <<http://InformationR.net/ir/8-3/paper154.html>>.

44. LINDBERG DA, HUMPHREYS BL. Medicine and health on the Internet: the good, the bad, and the ugly. *JAMA* 1998 Oct 21;280(15):1303-4.

45. ABAD GARCÍA MF, GONZÁLEZ TERUEL A, MARTÍNEZ CATALÁN C. La base de datos de recursos Web de la biblioteca médica virtual del COMV. *Hipertext.net* [serial online]. 2004 (2). [cited 26 Jul 2004]. <<http://www.hipertext.net/Web/page209.htm>>.

46. DISCONTINUING BIOSITES. Latitudes [serial online]. 2003 Sep/Oct;12(5). [cited 14 Mar 2005]. <<http://nnlm.gov/psr/lat/v12n5/biosites.html>>.

*Received August 2004; accepted April 2005*